



UNIVERSITY *of*
WEST FLORIDA

Introducing UWF-ZeekData: Network Datasets Based on the MITRE ATT&CK Framework

Dustin Mink, Sikha Bagui, & Subhash Bagui
University of West Florida, USA

- Introduction
 - Datasets, Cyber Range, and Big Data Platform
 - Cybersecurity Classes
 - Correlation
 - Questions
-

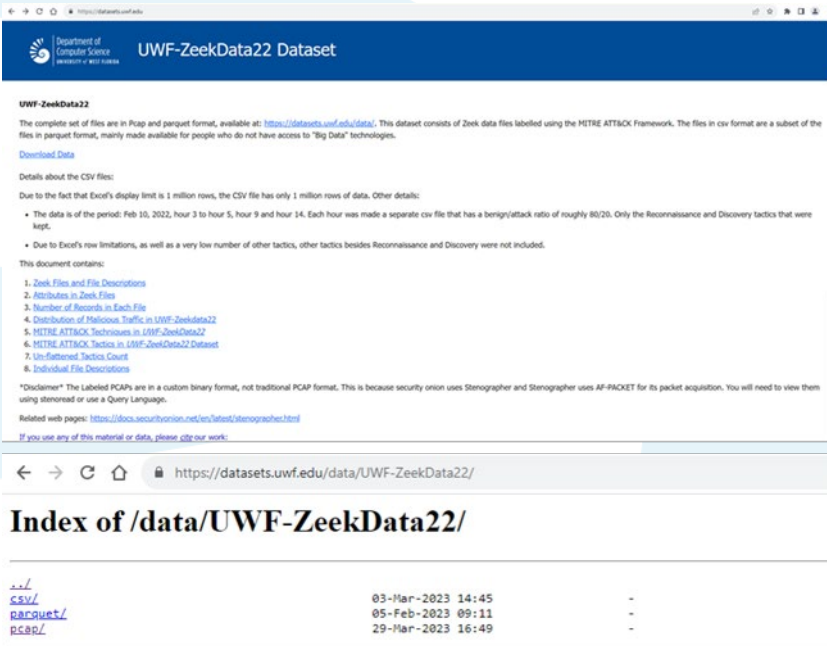
- Review major network intrusion detection dataset
- UWF-ZeekData22 is the first network intrusion detection dataset of Zeek logs labelled with the MITRE ATT&CK framework



UNIVERSITY *of*
WEST FLORIDA

**Datasets, Cyber Range, and
Big Data Platform**

- Publically available web site
- Zeek and MITRE ATT&CK in CSV and Parquet formats
- Raw network traffic in PCAP



The screenshot shows the 'UWF-ZeekData22 Dataset' page. It includes a 'Download Data' link, details about the CSV files (1 million rows, 1 million rows of data), and a list of contents. The page also features a disclaimer and related web pages.

UWF-ZeekData22

The complete set of files are in pcap and parquet format, available at: <https://datasets.uwf.edu/data/>. This dataset consists of Zeek data files labelled using the MITRE ATT&CK Framework. The files in csv format are a subset of the files in parquet format, mainly made available for people who do not have access to "Big Data" technologies.

[Download Data](#)

Details about the CSV files:

Due to the fact that Excel's display limit is 1 million rows, the CSV file has only 1 million rows of data. Other details:

- The data is of the period: Feb 10, 2022, hour 3 to hour 5, hour 9 and hour 14. Each hour was made a separate csv file that has a benign/attack ratio of roughly 80/20. Only the Reconnaissance and Discovery tactics that were kept.
- Due to Excel's row limitations, as well as a very low number of other tactics, other tactics besides Reconnaissance and Discovery were not included.

This document contains:

1. Zeek Files and File Descriptions
2. Attributes in Zeek Files
3. Number of Records in Each File
4. Distribution of Malicious Traffic in UWF-ZeekData22
5. MITRE ATT&CK Techniques in UWF-ZeekData22
6. MITRE ATT&CK Tactics in UWF-ZeekData22 Dataset
7. Un-filtered Tactics Count
8. Individual File Descriptions

Disclaimer The Labeled PCAPs are in a custom binary format, not traditional PCAP format. This is because security orion uses Stenographer and Stenographer uses AF-PACKET for its packet acquisition. You will need to view them using stenoextract or use a Query Language.

Related web pages: <https://docs.securityorion.net/en/latest/stenographer.html>

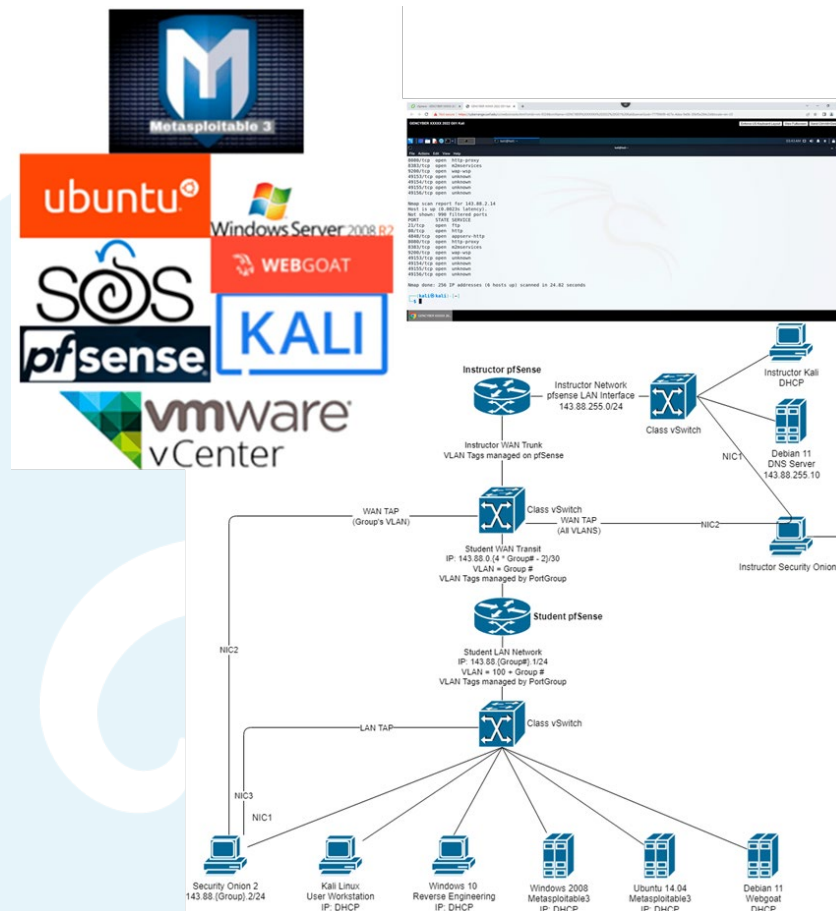
If you use any of this material or data, please cite our work:

<https://datasets.uwf.edu/data/UWF-ZeekData22/>

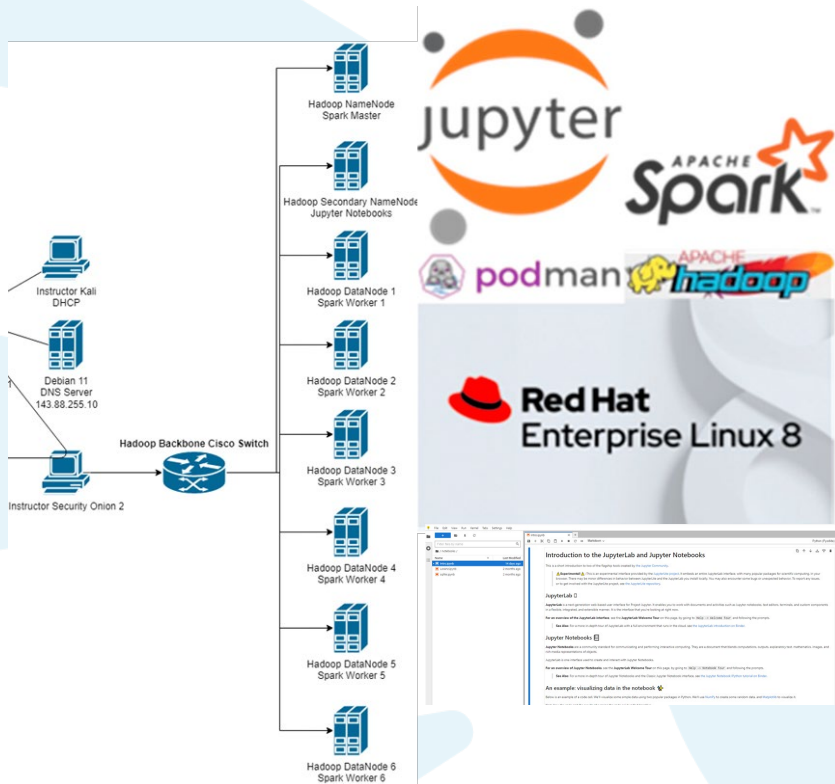
Index of /data/UWF-ZeekData22/

..	03-Mar-2023 14:45	-
csv/	05-Feb-2023 09:11	-
parquet/	29-Mar-2023 16:49	-
PCAP/		

- Supermicro X9DRE-TF+/X9DR7-TF+ (24CPU 128GB RAM 20TB HDD)
- Dell PowerEdge R740 (48CPU 768 GB RAM 6.74TB SSD)
- Dell PowerEdge R750 (128CPU 1TB RAM 12.43TB SSD)
- VMware vCenter
- VMware PowerCLI
- Kali, Security Onion, Pfsense, Metasploit 3 (Windows/Ubuntu), WebGoat
- Ethical Hacking and Penetration Testing and Cyber War Gaming



- (×2) Dell PowerEdge R730 (20CPU128GB RAM 4TB HDD)
- (×6) Dell PowerEdge R730xd (20CPU 128GB RAM 48TB HDD)
- Hadoop, Spark, Jupyter Notebooks
- Introduction to Big Data Analytics and Advanced Big Data Analytics





UNIVERSITY *of*
WEST FLORIDA

Cybersecurity Classes

- CAE-CD
Cybersecurity
Program
- Course Description
- Each student has their
own Kali VM
- Victims Windows
Metasploitable 3,
Ubuntu Metasploitable
3, and Ubuntu
WebGoat





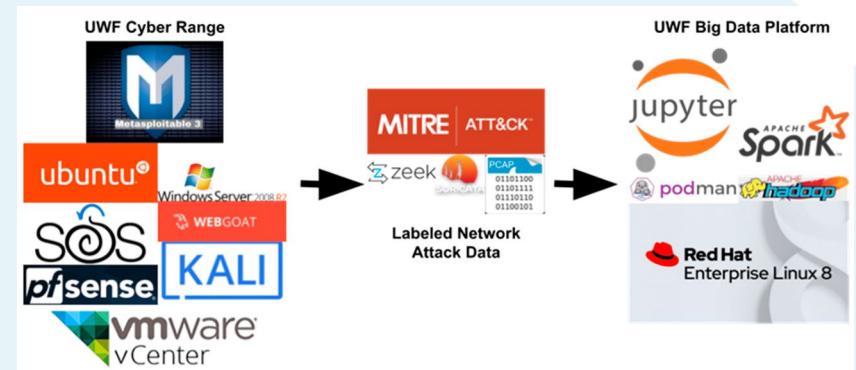
- CAE-CD Cybersecurity Program
- Course Description
- Pairs of Student team up for Capture the Flag
- Kali and Security Onion VMs



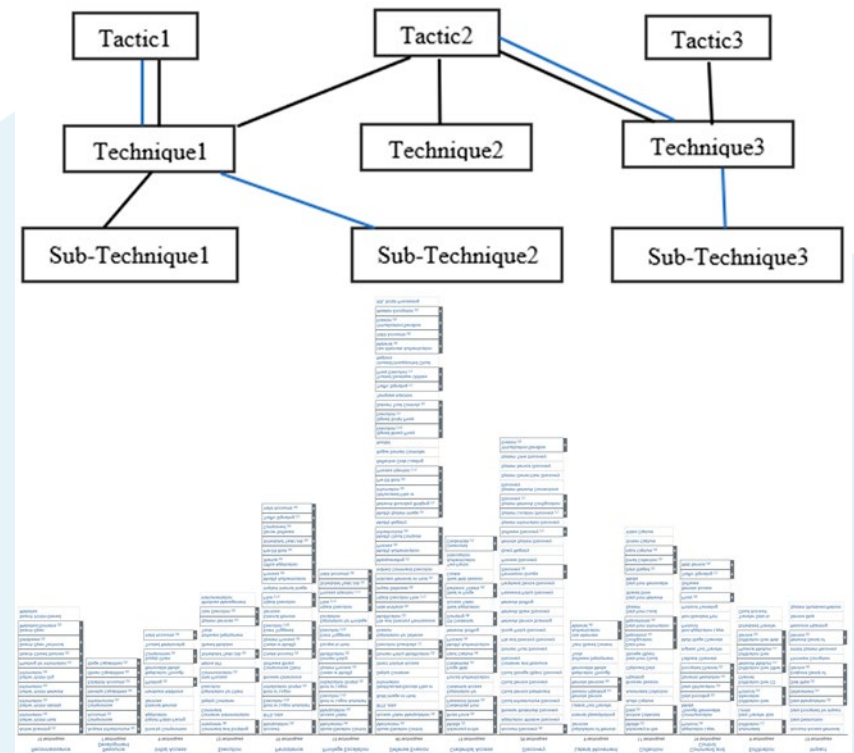
UNIVERSITY *of*
WEST FLORIDA

Correlation

- Instructure Security Onion 2 Collects Zeek and PCAPs
- CronTab runs Bash Script to transfer Zeek and PCAPs to HDFS
- Student enter metadata into Google Form
- End of the semester transfer mission logs to HDFS

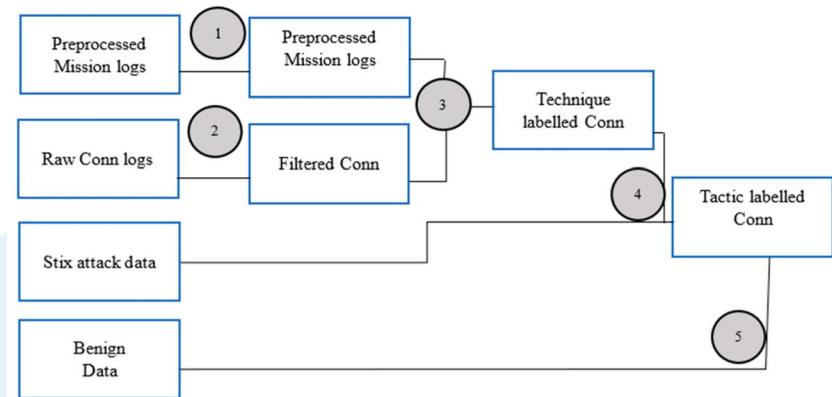


- MITRE ATT&CK cybersecurity industry standard
- UWF-ZeekData22, has 14 tactics, 191 techniques, and 358 sub-techniques



Correlate Zeek and Mission Logs

- Zeek Log
- Mission Log
- Start datetime, stop datetime, src ip, dest ip, src port, and dest port



```

#separator \x09
#set_separator ,
#empty_field (empty)
#unset_field -
#path conn
#open 2019-08-12-16-26-16
#fields ts uid id.orig_h id.orig_p id.resp_h
#duration uid id.orig_h id.orig_p id.resp_h
#missed_bytes history orig_pkts orig_ip_bytes orig_ip_bytes resp_pkts
nts
#types time string addr port addr port enum str
nt string bool bool count string count count
1517336042.090842 CBgZ5849tCIsWiMH39 10.55.182.100 142
- 3.000158 0 0 S0 - - 0
0 (empty)
1517336042.279652 CAcrIoTIOpfpNAS13 192.168.88.2 556
dns 0.069982 61 81 SF - - 0
109 (empty)

```

11183-CNT4416-202301 Mission Log

dmikguel.edu Switch account

* Indicates required question

Email *

Your email

What is your group number?

Your answer

What is the ID (e.g., T1548) of the Mine ATT&CK Technique (i.e., <https://attack.mitre.org/techniques/mitre/enterprise/1548>)? (Link to an external site.) your group used (i.e., submit one survey per Mine ATT&CK Technique. One attack can have one or more Mine ATT&CK Techniques)?

Your answer

What is the source IP of the attack (e.g., 143.88.0.1 or 143.88.0.0/24)?

Your answer

What is the source port of the attack (e.g., 8888)?

- Zeek Connection Log
- MITRE ATT&CK Tactic

```
>>> df_conn.printSchema()
root
 |-- resp_pkts: integer (nullable = true)
 |-- mitre_attack: string (nullable = true)
 |-- service: string (nullable = true)
 |-- orig_ip_bytes: integer (nullable = true)
 |-- local_resp: boolean (nullable = true)
 |-- missed_bytes: integer (nullable = true)
 |-- proto: string (nullable = true)
 |-- duration: double (nullable = true)
 |-- conn_state: string (nullable = true)
 |-- dest_ip_zeek: string (nullable = true)
 |-- orig_pkts: integer (nullable = true)
 |-- community_id: string (nullable = true)
 |-- resp_ip_bytes: integer (nullable = true)
 |-- dest_port_zeek: integer (nullable = true)
 |-- orig_bytes: integer (nullable = true)
 |-- local_orig: boolean (nullable = true)
 |-- datetime: timestamp (nullable = true)
 |-- history: string (nullable = true)
 |-- resp_bytes: integer (nullable = true)
 |-- uid: string (nullable = true)
 |-- src_port_zeek: integer (nullable = true)
 |-- ts: double (nullable = true)
 |-- src_ip_zeek: string (nullable = true)
```

- UWF-ZeekData22 is the first network intrusion detection dataset of Zeek logs labelled with the MITRE ATT&CK framework

Acknowledgements

- The Research was supported by 2021 NCAE-C-002: Cyber Research Innovation Grant Program, Grant Number: H98230-21-1-0170



UNIVERSITY *of*
WEST FLORIDA

Questions?

Data available at: <https://datasets.uwf.edu>